The EMBO Journal vol.14 no.1 pp.151–158, 1995

The identity of the base following the stop codon determines the efficiency of *in vivo* translational termination in *Escherichia coli*

Elizabeth S.Poole, Chris M.Brown[1] and Warren P.Tate[2]

Department of Biochemistry and Centre for Gene Research, University of Otago, Dunedin, New Zealand
[1]Present address: Plant Pathology Department, 351 Bessey Hall, Iowa State University, Ames, IA 50011, USA

[2]Corresponding author

Communicated by A.-L.Haenni

A statistical analysis of >2000 *Escherichia coli* genes suggested that the base following the translational stop codon might be an important feature of the signal for termination. The strengths of each of 12 possible 'four base stop signals' (UAAN, UGAN and UAGN) were tested in an *in vivo* termination assay that measured termination efficiency by its direct competition with frameshifting. Termination efficiencies varied significantly depending on both the stop codon and the fourth base, ranging from 80 (UAAU) to 7% (UGAC). For both the UAAN and UGAN series, the fourth base hierarchy was U > G > A ~ C. UAG stop codons, which are used rarely in *E.coli*, showed efficiencies comparable with UAAN and UGAN, but differed in that the hierarchy of the fourth base was G > U ~ A > C. The rate of release factor selection varied 30-fold at UGAN stop signals, and 10-fold for both the UAAN and UAGN series; it correlated well with the frequency with which the different UAAN and UGAN signals are found at natural termination sites. The results suggest that the identity of the base following the stop codon determines the efficiency of translational termination in *E.coli*. They also provide a rationale for the use of the strong UAAU signal in highly expressed genes and for the occurrence of the weaker UGAC signal at several recoding sites.

Key words: frameshifting/highly expressed genes/release factors/stop codon/translational termination

Introduction

The termination of protein synthesis involves an intimate interaction between the ribosome, mRNA and proteins called release factors (RFs; Craigen *et al.*, 1990; Tate and Brown, 1992). With the elucidation of the genetic code, protein synthesis termination signals in the mRNA were defined as simple triplet codons (UAA, UAG and UGA; Crick, 1966). In prokaryotes there are two RFs that each recognize two of the three stop codons: RF-1 recognizes UAG and UAA; RF-2 recognizes UGA and UAA (Scolnik *et al.*, 1968). However, early suppression studies suggested that there may be a larger context than three nucleotides specifying termination (Salser, 1969; Salser *et al.*, 1969; Fluck *et al.*, 1977; Bossi and Roth, 1980; Bossi, 1983). There is now growing evidence that the stop signal is recognized directly by the RF proteins, and so the recognition determinant need not be restricted to simply the triplet codon (Brown *et al.*, 1990; C.Brown and W.Tate, manuscript in preparation).

The two bases immediately 3′ to UGA stop codons influence UGA suppression, and it has been speculated that these bases affect termination (Buckingham *et al.*, 1990). Indeed, Martin *et al.* (1988) demonstrated that RF-1 and RF-2 had different relative activities depending on the site of UAA codons within a gene, with an inverse correlation between the relative activity of the factors and the efficiency of suppression. Stormo *et al.* (1986) had earlier proposed that a significant factor in the effect of the two 3′ bases on nonsense suppression resulted from 3′ 'dangling bases' stabilizing the codon–anticodon interaction. It is likely that these context effects on suppression result from a combination of tRNA- and RF-mediated causes (Yarus and Curran, 1992).

Further evidence in support of the idea that the stop signal might extend beyond the three nucleotides has been provided by the discovery of new alternative coding or recoding events which occur at a small number of specific stop sites (Gesteland *et al.*, 1992). In these positions the stop signal can signify either a complete stop in protein synthesis or a pause for more specialized recoding events beyond the usual constraints of the genetic code (Tate and Brown, 1992). This pause may allow a diverse array of possible competing events at the stop signal to take place, including specific amino acid incorporation (Böck *et al.*, 1991), suppressor tRNA readthrough (Hatfield *et al.*, 1990) or frameshifting (Atkins *et al.*, 1990).

If the signal for translational termination were larger than a triplet codon, then statistical evidence may be found in the sequence surrounding the stop codon, since the additional feature might be constrained during evolution. Indeed, statistical analyses of the sequences around the stop codons of *Escherichia coli* genes have shown localized non-randomness (Brown *et al.*, 1990; Arkov *et al.*, 1993). In particular, there was a significant bias in the nucleotide immediately following the triplet stop codon (Brown *et al.*, 1990). These findings suggested not only that an extended stop signal might be recognized by RF proteins, but also that the fourth nucleotide might influence the efficiency of the signal. However, these statistical analyses did not take into account the frequency bias of some tetranucleotides in the *E.coli* genome; for example, the tetranucleotide CTAG is a rare sequence throughout the genome (Phillips *et al.*, 1987). Despite this, the predictions of the statistical analysis were supported by studies of the rate of RF-1 selection at the four UAGN termination signals in *E.coli* (Pedersen and Curran, 1991).

In this study we have tested experimentally the hypo-

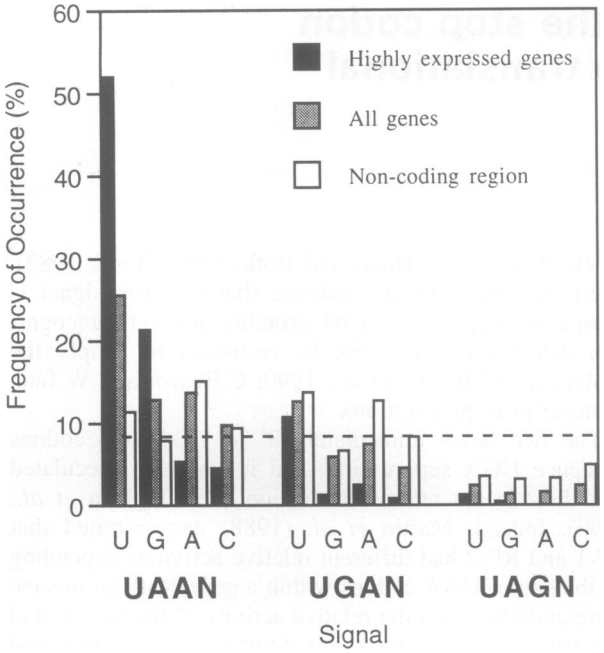

Fig. 1. The relative frequency of occurrence of tetranucleotide stop signals in 2455 *E.coli* genes. The dotted line shows the random expectation of each tetramer. Each fourth base is shown above the termination signal.

thesis that the base following the stop codon influences the efficiency by which the stop codon is decoded. An *in vivo* termination assay was used in *E.coli* that measures termination efficiency in competition with an alternative recoding event to test the strength of each tetranucleotide stop signal. Our data show that the 12 possible four base signals differ greatly in how effectively they specify termination against the competing event, implying that the fourth base is a significant determinant of the strength of the stop signal.

Results

The relative occurrence of tetranucleotide stop signals in E.coli genes

Our original study in 1990 analysed the stop codon context in 862 *E.coli* genes (Brown *et al.*, 1990). A total of 12 stop signals are generated if a fourth base is included and we have now analysed the use of these in 2455 *E.coli* genes currently available for study (Brown *et al.*, 1993a, 1994). This dataset should be representative since it is likely to encompass about half of the genes in *E.coli*. The subset of these genes, which had a sense codon bias most similar to highly expressed genes, has been analysed separately. The rationale for considering this set of genes was that the codon bias indicates an evolutionary pressure for the optimization of translation elongation and, therefore, they might also be expected to use optimal termination signals.

Since the G + C content of the *E.coli* genome is close to 50%, ~8.3% of each of the 12 tetramers would be expected randomly (Figure 1, dotted line). However, the occurrence of these tetranucleotides in non-coding regions is clearly not random. UAAA, UGAU and UGAA are over-represented, for example, whereas the four UAGN signals, and to a lesser extent UGAG and UAAG, are under-represented. These frequencies reflect non-random utilization of di-, tri- and tetranucleotides in the whole genome (Figure 1, open bars). The occurrence of these tetramers at stop signals has been compared with their occurrence in the non-coding regions and thus permitted identification of bias specific to stop signals (Figure 1). For all *E.coli* genes, the UAAU and UAAG were much more frequent at stop codons than expected from their occurrence in the non-coding regions. UAAC occurs at a frequency slightly more than expected, whereas most of the other nine signals have apparently been selected against (hatched bars). This pattern is even more marked in the highly expressed genes (solid bars) with most of this subset using UAAU or UAAG, in agreement with our earlier statistical analyses (Tate and Brown, 1992). Therefore, in this subset the apparent hierarchy for termination is UAAU > UAAG > UAAA ~ UAAC.

For the UGAN series, UGAU is not greatly selected against, whereas the others are rarely used. In particular, UGAC is infrequently found at stop codons. Interestingly, this UGAC sequence is found as the signal at two sites where alternative recoding events occur in *E.coli*: the UGA used for selenocysteine incorporation in the formate dehydrogenase mRNA, and the UGA avoided in the +1 frameshift in the translation of the RF-2 protein (Craigen *et al.*, 1985; Zinoni *et al.*, 1986; Böck *et al.*, 1991). All of the UAGN tetranucleotides are selected against as stop signals, even when the low frequency of these sequences in non-coding regions of genes is taken into account.

The statistical analyses are provocative and we speculate that the *E.coli* RF proteins might recognize a tetranucleotide stop signal, with the fourth base determining the efficiency by which the signal is decoded by RFs. A rapid rate of stop signal selection would result in efficient termination, whereas a slower rate of selection might be less compatible with the required high rate of expression of some genes or might favour competing events. We have tested this tetranucleotide hypothesis in an *in vivo* assay to measure the relative termination efficiencies of all 12 four base signals.

The translational termination/frameshift assay system

Expression of RF-2 in *E.coli* requires a +1 frameshifting event to avoid an in-frame stop codon (Craigen *et al.*, 1985). The stop signal itself contributes to a pause in translation that is believed to enhance ribosomal slippage over a run of uracils on the mRNA immediately 5′ to the stop signal (Weiss *et al.*, 1987; Hatfield and Oroszlan, 1990). For frameshifting to occur, the elements 5′ of the stop codon promoting the event must compete with translational termination at the stop codon. *In vivo*, the two events compete almost equally with frameshifting efficiencies of 30–50% (Craigen and Caskey, 1986; Donly *et al.*, 1990). If the stop signal were altered in this system, then the degree of frameshifting should be influenced by the efficiency of the sequence to signal stop. Indeed, a change in either the concentration of RF-2 or its specific activity can change the efficiency of decoding the natural UGAC as stop, thereby affecting frameshifting over a 0–100% range (Craigen and Caskey, 1986; Kawakami *et al.*, 1988; Donly *et al.*, 1990).

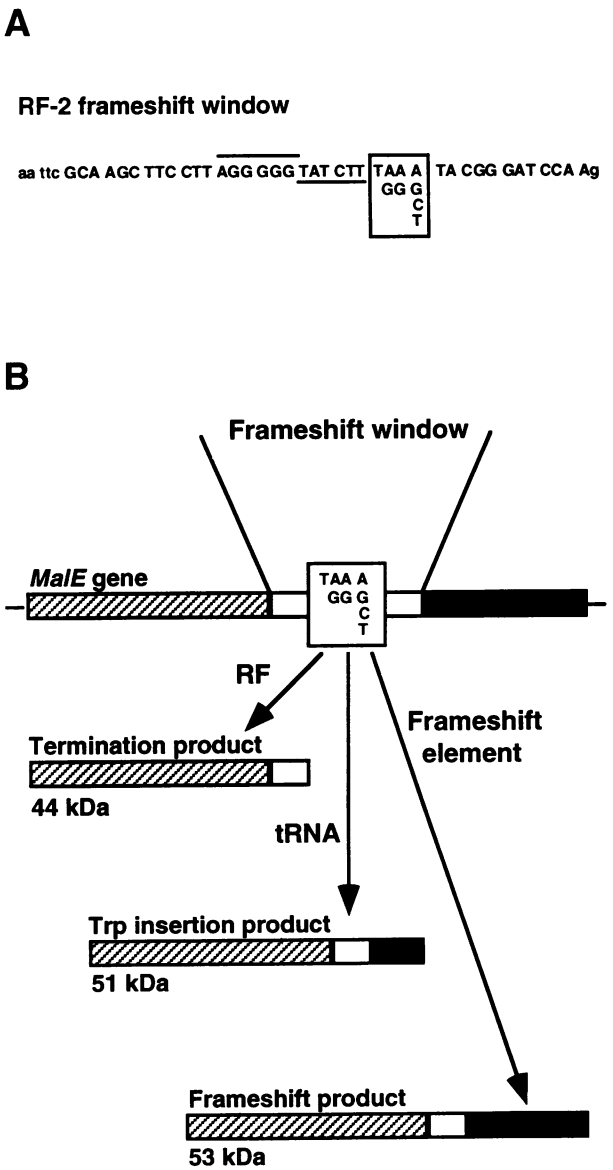

Fig. 2. The translational termination\frameshift assay system. (**A**) Sequence of the RF-2 frameshift 'window' containing the Shine–Dalgarno sequence (overlined) and the 'slippery run' (underscored) essential for frameshifting. The redundancies introduced at the stop codon site are shown (box). The restriction endonuclease sequences are shown in lowercase. (**B**) Construction of the *MalE* gene–RF-2 frameshift 'window' fusions and the possible protein products produced after expression. The molecular weight (kDa) of each protein product is shown. The mediators of the three events are shown beside the arrows.

We have used the RF-2 frameshift 'window' (Curran and Yarus, 1988) fused to the *MalE* gene in the plasmid pMAL™ to investigate the relative efficiencies of tetranucleotide stop signals *in vivo*. A diagram of the test system is shown in Figure 2. Sequences were constructed to contain the natural elements that promote the +1 frameshifting event, a Shine–Dalgarno sequence (AGG–GGG) that can base-pair to the 16S rRNA during ribosomal binding (Weiss *et al.*, 1988), spaced appropriately from a run of 'slippery' uracils immediately preceding the stop codon. At the stop codon site redundancies were introduced so that all 12 tetranucleotide stop signals (UAAN, UGAN and UAGN) would be represented. This method of construction also created the UGGN series which lacks a stop codon and instead contains the UGG tryptophan codon. Plasmid-encoded proteins expressed *in vivo* were analysed with an antibody to maltose binding protein (MBP) which will detect both the termination and frameshift products, or the product from tryptophan insertion and frameshift product in the case of the UGGN series.

Independent isolates of each clone were analysed for stop signal strength in three separate experiments. In addition, we analysed one series (UAGN) in minimal media (M9) as well as in rich media (LB) and demonstrated that the media did not significantly influence the rate of RF selection at a particular signal.

Influence of tetranucleotide signals on the efficiency of translational termination

Three protein products can result from translation of the mRNA constructs: (i) a 44 kDa fusion protein resulting from termination at the stop signal; (ii) a 53 kDa fusion protein produced from a +1 frameshifting event; and (iii) a 51 kDa readthrough protein formed from either amino acid incorporation at the stop site in response to a suppression event, or cognate aminoacyl-tRNA^{Trp} recognition of the UGG signal for that series (Figure 2). We have used a wild-type *E.coli* strain that does not contain a UAG suppressor tRNA to eliminate suppression as a third potential competing event with termination and frameshifting (Adamski *et al.*, 1993). At the frameshift site there will be competition between termination and frameshifting, such that the ratio of the amounts of the expressed fusion proteins should reflect the rate of RF selection. For the UGGN series, there will be competition between tryptophan incorporation and frameshifting that will reflect the rate of cognate aminoacyl-tRNA selection.

The ratio of the expressed protein products was determined for each construct. Figure 3A and B shows the results from the constructs containing the UAAN signals. Figure 3A, C and E represents the data from a single experiment on single clones from each of the 12 constructs. The means of several experiments with multiple isolates of each clone are represented in Figure 3B, D and F. The ratio of the termination product (44 kDa) to frameshift product (53 kDa) changes significantly when the fourth base of the signal is altered (Figure 3A). The termination efficiency varies 3-fold depending on the 3′ nucleotide in the order of efficiency UAAU > UAAG > UAAA ~ UAAC (Figure 3B). The stop signals competing best with frameshifting were those that were most commonly used for translational termination in *E.coli* genes (Figure 1). UAAU is the stop signal in 52% of our set of highly expressed genes in *E.coli*.

The results for the UGAN series follow a similar pattern to the UAAN series in the hierarchy of termination efficiency: UGAU > UGAG > UGAA > UGAC (Figure 3C and D). However, in this case the efficiency varies more widely. UGAC is clearly the poorest termination signal of any of the stop signals. Conversely, UGAU is a highly effective termination signal, similar in this respect to UAAU, and therefore would be effective in highly expressed genes requiring efficient termination. Indeed, UGAU is the stop signal in 11% of our set of highly expressed genes in *E.coli*.

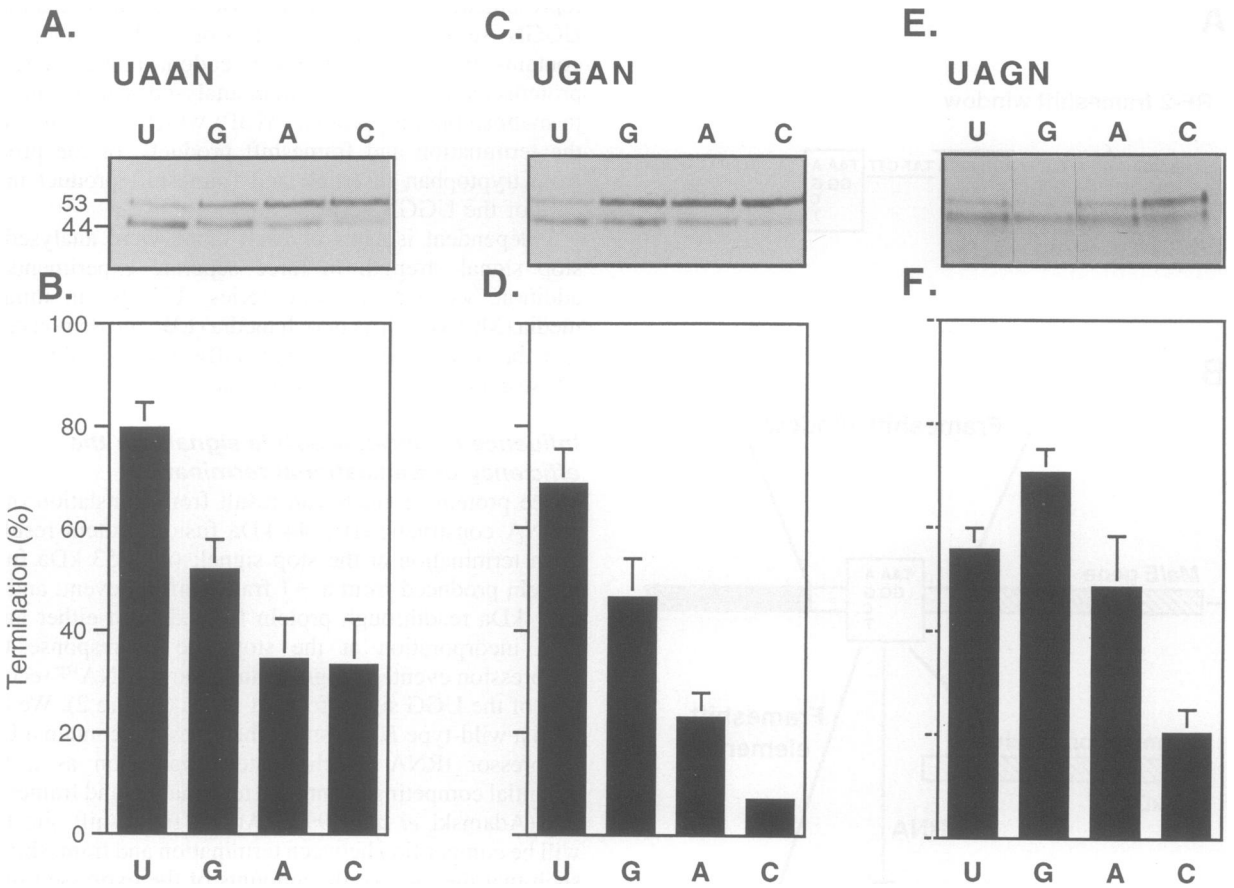

Fig. 3. Influence of tetranucleotide stop signals on the efficiency of translational termination. (**A** and **B**) UAAN signals, (**C** and **D**) UGAN signals and (**E** and **F**) UAGN signals. The protein products (A, C and E) and the termination efficiency for each signal, expressed as a percentage (B, D and F), are shown. The molecular weights of the termination (44 kDa) and frameshift products (53 kDa) are shown. The frameshift product for UAGG (10%) (E) for this experiment was atypical and, although visible, was faint. (This clone, along with other isolates, in 12 out of 14 experiments for UAGG constructs gave ~30% frameshifting.) The fourth base is shown in each lane. The standard deviation is shown and was calculated using the data derived from at least three experiments and in the majority of cases multiple isolates of each clone.

UAGN tetranucleotide signals might be predicted to be poor termination signals from the statistical analysis alone. However, in this *in vivo* system UAGN stop signals showed efficiencies comparable with those of UAAN and UGAN signals (Figure 3E and F) and varied in the order UAGG > UAGU ~ UAGA > UAGC.

It was of interest to examine the effect of the fourth base on the rate of selection of a cognate aminoacyl-tRNA^{Trp} at a UGGN signal, compared with the rate of selection of RF at stop signals. In this case, the competing events are the rate of selection of aminoacyl-tRNA^{Trp} (normal amino acid incorporation) versus frameshifting. Readthrough efficiencies were similar for UGGA, UGGG and UGGU, but significantly lower for UGGC. This result most probably reflects base stacking, a context effect where a 3′ unpaired nucleotide stabilizes the codon–anticodon helix (Yarus and Curran, 1992).

Rates of selection of RFs

The rates of RF selection relative to frameshifting for each stop signal series were estimated from termination and frameshifting percentages within each experiment. RF-1 is selected at UAAN and UAGN signals and RF-2 at UAAN and UGAN signals, and the selection rate is compared for each of the tetranucleotide signals (Table I). The rate of RF selection when U follows UAA or UGA is significantly faster than when other nucleotides are in this position. The most striking feature of the data is that the combined rate of RF-1 and RF-2 selection at UAAU stop signals (5.2) is almost twice as fast as at any other stop signal and provides a reason for the bias towards the use of this signal in highly expressed genes. The RF-1 selection rate at UAGG (3.0) is significantly faster than

Table I. Relative rates of RF and aminoacyl-tRNA^{Trp} selection at each tetranucleotide signal

	U	G	A	C
Rate of RF selection				
UAAN	5.2	1.1	0.6	0.5
UGAN	2.8	1.0	0.3	0.1
UAGN	1.3	3.0	1.1	0.3
Rate of selection of aminoacyl-tRNA^{Trp}				
UGGN	1.5	2.7	3.2	0.4

The rate of selection is the rate of RF or aminoacyl-tRNA selection relative to the rate of frameshifting. These values are calculated using an equation derived by Pedersen and Curran (1991). The fourth base is shown at the head of the Table.

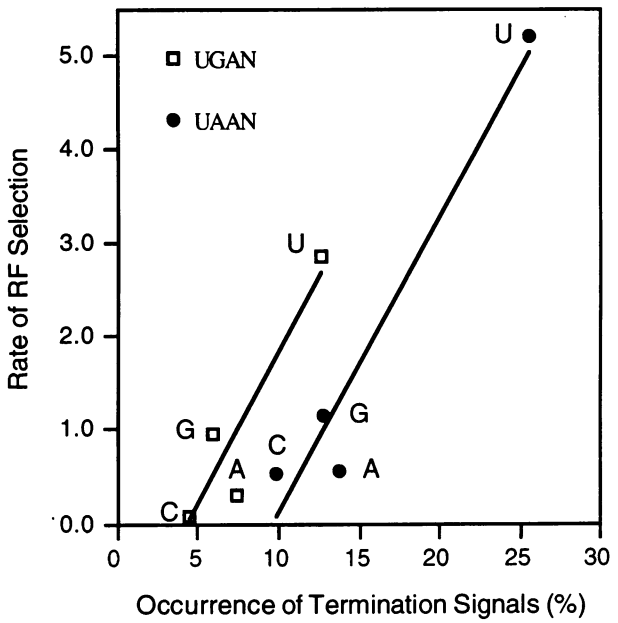

Fig. 4. Relative rates of RF selection at UGAN signals (RF-1) and UAAN signals (RF-1 and RF-2) as a function of the occurrence of each signal in *E.coli*. The values for the relative rates of RF selection are shown in Table I. The numbers of each termination signal are expressed as a percentage of the total number of signals and were derived from the statistical analysis (Figure 1). The linear correlation coefficient (r) is 0.85 for UGAN signals and 0.95 for UAAN signals.

rates for the other 3′ nucleotides in that series and is almost identical to that of RF-2 at UGAU (2.8). The rate of RF-2 selection for UGAN signals varied nearly 30-fold due to the weakness of UGAC as a stop signal, in contrast to the 10-fold variation in RF selection rate for the UAAN and UAGN series. The rate of selection of a cognate aminoacyl-tRNA^{Trp} at UGGN signals was of the same magnitude as the selection rates of either RF at stop signals.

The identity of the fourth base of naturally occurring tetranucleotide stop signals was non-random (Figure 1). We used a linear regression analysis to determine whether the relative rates of RF-1 and RF-2 selection for UAAN contexts and RF-2 for UGAN contexts (the stop signals used almost exclusively in *E.coli*) correlated with the usage bias (Figure 4). Both stop signal contexts correlated strongly, with correlation coefficients of $r = 0.95$ for UAAN contexts and $r = 0.85$ for UGAN contexts. This correlation strongly indicates that the stop signals used in *E.coli* have evolved to promote RF selection by the ribosome. A linear regression analysis of the rate of RF-1 selection at UAGN signals did not correlate with usage (data not shown). This lack of correlation further suggests that a cause exists for the scarcity of UAGN signals in *E.coli* that is not related to the translational termination mechanism itself.

Discussion

The frameshift/termination assay system measures all outcomes of translation

Frameshifting and termination at the natural RF-2 +1 frameshift site (UGAC) are in direct competition (Craigen *et al.*, 1985; Craigen and Caskey, 1986; Curran and Yarus, 1988, 1989; Donly *et al.*, 1990; Weiss *et al.*, 1990; Adamski *et al.*, 1993; Curran, 1993). This well-characterized competition provides an ideal test system to compare the strength of different tetranucleotide stop signals when faced with competition from an alternative event.

The RF-2 frameshift sequence has been used for *in vivo* experiments to examine autoregulation of RF-2 mRNA (Curran and Yarus, 1988), rates of aminoacyl-tRNA selection (Curran and Yarus, 1989), rates of suppressor tRNA selection compared with rates of RF-1 selection (Pedersen and Curran, 1991) and the effects of tRNA:message stability on frameshift frequency (Curran, 1993). All of these studies used constructions in which the frameshift window was fused to a *lacZ* reporter gene and a +1 frameshift was required for the synthesis of β-galactosidase. One limitation of this model system is that only one of the two (or three) events at the site can be measured directly, namely the frameshift event, and for comparisons between samples some normalization procedure should be considered. Our system obviates this requirement at equivalent levels of expression by allowing direct measurement of all possible outcomes of translation from each mRNA assay: termination, frameshifting or read-through.

In characterizing the RF-2 frameshift region, Weiss *et al.* (1990) changed many sites in the frameshift window by mutagenesis and compared β-galactosidase activities as a measure of relative frameshifting efficiency. In some of these constructs, the fourth base was altered along with other bases. If one now interprets these relevant data according to our hypothesis, then despite the changes additional to the fourth base and the lack of normalization of mRNA levels, the data are in excellent agreement with our results.

The majority of highly expressed genes uses very efficient UAAU or UAAG stop signals

The statistical analysis found that a total of 62% of *E.coli* genes used UAAN stop signals. The most striking finding of this analysis was that UAAU and UAAG were found significantly more frequently as stop signals than in non-coding regions. These two signals specified termination in 73% of highly expressed genes. These results corroborated the earlier statistical analysis that suggested a strong preference for UAAU-containing signals at the end of highly expressed genes (Brown *et al.*, 1990). In agreement with the statistical analyses, the experimental results found that UAAU was an extremely efficient termination signal, indeed the most efficient of all the 12 possible tetranucleotide signals. In a previous study of natural suppression in *E.coli*, suppression efficiencies at UAA were between 10^{-3} and 10^{-5}, the lowest of all the stop codons. Cells that contained UAA suppressors grew slowly (Rydén and Isaksson, 1984; Eggertsson and Söll, 1988). The abundance of UAAN signals suggests that it is important that suppression at UAAN signals be inefficient, as read-through in the translation of highly expressed genes could severely compromise bacterial viability.

UGAN stop signals vary widely in termination efficiency

The statistical analysis showed that nearly 11% of highly expressed genes used UGAU as the termination signal,

the third most common stop signal after UAAU and UAAG. This indication that UGAU could be an efficient stop signal was corroborated by the experimental results, where termination efficiency approached that of UAAU and was better than UAAG. Consistent with this finding, an *in vivo* study of tRNA suppression efficiency at UGAN codon contexts found that suppression was least efficient at UGAU (Kopelowitz *et al.*, 1992). If UGAU stop signals are effective, then it might be expected that this signal would be used at a greater frequency in highly expressed genes. A probable reason for the selection against UGA-containing signals could be that some UGA codons are relatively easily suppressed. UGA codons can be read by $tRNA^{Trp}$ *in vitro* (Hirsh and Gold, 1971) and UGA suppression was found to occur in wild-type *E.coli* strains at a level of $10^{-2}-10^{-4}$ (Parker, 1989). This suppression may be greatest at UGA codons followed by A or G (Kopelowitz *et al.*, 1992). Our *in vivo* results showed that UGAC, when faced with a competing event, was an extremely poor termination signal. The wide range of termination efficiencies for UGAN signals correlated with the occurrence of these signals at natural termination sites, and provided a reason for finding the poor signal, UGAC, at sites of certain recoding events. In this context, UGA is recognized by a specific aminoacyl-tRNA which incorporates selenocysteine into the protein formate dehydrogenase (Böck *et al.*, 1991) and is the natural signal for the +1 frameshift in the autoregulation of expression of RF-2 in *E.coli* (Craigen and Caskey, 1986; Donly *et al.*, 1990).

RF-1 selection at UAGN stop signals is efficient, although they are rarely used as stop signals

The rate of RF-1 selection at UAGN stop signals showed efficiencies comparable with RF selection rates at either UAAN or UGAN stop signals, despite the finding from the statistical analysis that UAGN signals are rarely used for termination (Brown *et al.*, 1990). Although TAGN signals are relatively scarce in the *E.coli* genome, there must be an additional bias against their use as termination signals since the *in vivo* assay showed that RF-1 recognition of UAGN is not poor. Notably, a significant number (30%) of wild-type *E.coli* strains contain UAG suppressor tRNAs (Marshall and Levy, 1980) and UAG-specified translational termination would be inefficient in such strains. However, it appears that these suppression events are tolerated with no apparent selective disadvantage. UAGN stop signals, under certain conditions, might be mistakenly altered to a sense codon by the mechanism for *vsr*-initiated DNA mismatch repair. The vsr gene product is an endonuclease which nicks one strand at a T:G mismatch, at the T in CTRG or TRGG sequences. This mechanism may be the cause of the low frequency of TAGN sequences in the *E.coli* genome (McClelland and Bhagwat, 1992). While stop signals converted to sense by either of these mechanisms might be tolerated if only a small number of the less essential genes used them, it would compromise viability if a large proportion of genes were affected. *E.coli* strains that used UAAN or UGAN signals for translational termination for the majority of genes might have had a selective advantage in the course of evolution.

The RF-2 gene +1 frameshift sequence has been used to measure frameshift β-galactosidase units to compare the rates of suppressor tRNA and RF-1 selection at UAGN signals (Pedersen and Curran, 1991). In our study, the relative rates of RF-1 selection at UAGN signals were in the hierarchy G > U > A > C, the same hierarchy as found in this earlier study before, but not after, they normalized for mRNA levels. The experiments described in our study measure both termination and frameshifting, the total outcome of translation from each mRNA molecule rather than just the frameshifting event, and therefore normalization for mRNA levels should not be required. Moreover, the amounts of the products (frameshift plus termination) obtained across the series were very similar.

The rate of RF selection varies at different tetranucleotide stop signals and allows for control of translational termination

The RF selection rate varied widely within each stop signal series and depended on the identity of the fourth base, with that at UAAU the fastest of all the signals. UAA stop codons are recognized by both RF-1 and RF-2 and the effective concentration of the decoding factor would therefore be higher at UAAN signals. The experimental results show this dramatically in the case of UAAU where the RF-1 and RF-2 selection rates were nearly twice that of RF-2 selection at UGAU and RF-1 selection at UAGG. This is consistent with the experimental evidence of a strict maintenance of the ratio of the two RFs over all growth rates and does not support the possibility that one of the RFs might have been used preferentially at high growth rates when there is a requirement for high expression (Adamski *et al.*, 1994). However, for the other UAAN signals which do not show this effect, one factor may be favoured. A temperature-sensitive mutant of RF-1 allowed differential misreading at UAG contexts over UAA (or as expected UGA), but varied from one UAA context to another, suggesting that the relative contributions of RF-1 and RF-2 to the decoding of these UAA signals were different (Rydén and Isaksson, 1984). Moreover, previously RF-1 was found to be favoured at UAA codons efficiently suppressed (UAAA/G) and RF-2 at sites poorly suppressed (UAAU/C) (Martin *et al.*, 1988).

Translation of highly expressed genes at fast growth rates requires efficient stop signals. At peak growth rates the protein synthesis apparatus dominates the cell. If translational termination was particularly slow, then ribosomes could stall behind the stop signal and not be released quickly to initiate another cycle of translation. Rapid RF selection at stop signals supports rapid ribosome recycling and fast translation. The results from this study showed that at certain stop signals (UAAU, UGAU and UAAG) the rate of RF selection was extremely rapid, and that these were also the signals most used at the end of highly expressed genes. Conversely, the rate of RF selection at UGAC, for example, was relatively poor and could promote ribosome stalling, providing the opportunity for alternative recoding events to occur. Experiments are in progress to investigate the degree and effect of ribosome stacking behind natural stop signals at both the end of genes and at recoding sites.

Translational termination in *E.coli* has evolved to balance efficiency and processivity (Jørgensen and

Kurland, 1990). The results described here suggest that this is achieved through the different affinities of the RFs for stop signals and the extreme stop signal bias in highly expressed genes. The rate of selection of RF was dependent on the identity of the fourth base of the tetranucleotide signal. Although it had been suggested that RFs may not interact with mRNA directly (Prescott *et al.*, 1991), there is now growing evidence from crosslinking experiments of a close interaction between the RF and stop signal through hydrogen bonding (Tate *et al.*, 1990; C.Brown and W.Tate, manuscript in preparation). Additional hydrogen bonding to the common imino donor and keto acceptor positions of either U or G in the fourth position of the stop signal, as suggested by Yarus and Curran (1992), could explain the results observed *in vivo* in this study. However, the actual mechanism by which the fourth base modulates the efficiency of the stop signal is yet to be determined. This base may be recognized directly by the RF, or it may influence the conformation of the termination codon and, therefore, its interaction with RF.

Alternatively, the bases U and G in the fourth position may interact better with the decoding region of the 16S rRNA, providing a higher affinity substrate for the release factor. Indeed, an RNA analogue of this decoding region from the 30S ribosomal subunit can interact with antibiotics and RNA ligands, simulating normal function (Purohit and Stern, 1994). A putative site of termination signal interaction in the decoding region of the 16S rRNA (1405–1409) has been suggested by a crosslinking study (Tate *et al.*, 1990). The bases (1406–1408) could pair with all three termination codons, but G1405 would be expected to interact with fourth base C or U rather than A or G (Brown *et al.*, 1993b). This pyrimidine/purine split, however, does not correlate with the finding in our current study that U or G was more effective than C or A in influencing termination efficiency.

We believe the four base concept may be applicable to most organisms. Using an *in vitro* termination assay based on mammalian components, we have obtained evidence for an influence of the fourth base on eukaryotic RF (eRF) decoding of the termination signal which, while different from the order derived for *E.coli* in this study, is consistent with that predicted from the statistical analysis of >5000 mammalian genes (K.K.McCaughan, C.M.Brown, M.E.Dalphin, M.J.Berry and W.P.Tate, unpublished data). The generality of the concept of a four base stop signal is currently being rigorously examined.

Materials and methods

Materials

The pMAL™-c2 plasmid and antibody to MBP were purchased from New England Biolabs, restriction endonucleases from Amersham or New England Biolabs, and T4 DNA ligase from Boehringer. [γ-^{32}P]dATP was purchased from Amersham. Deoxyoligonucleotides were made using an Applied Biosystems 380B DNA synthesizer, and cloned DNA was sequenced using a 373A ABI sequencer. Plasmids were prepared using a Magic™ Minipreps DNA purification system (Promega) and were electroporated into bacterial cells using an Electro Cell Manipulator© 600 (BTX). A Mini-PROTEAN II electrophoresis cell (Bio-Rad) and Mini Trans-Blot electrophoretic transfer cell (Bio-Rad) were used for gel electrophoresis and protein transfer, respectively. A GS-670 imaging densitometer (Bio-Rad) was used for laser densitometry.

Computer sequence analysis

The 2455 *E.coli* sequences analysed were obtained from the TransTerm database (release of June 1994; Brown *et al.*, 1994). The termination codon contexts for the genes analysed and summary statistics are available by E-mail or anonymous ftp from the EMBL server in the TransTerm directory (Rice *et al.*, 1993). These termination codon contexts were extracted as described in the documentation for the database. Briefly, the information in the feature tables of GenBank entries (flat file format release 82, April 1994) was used to extract the contexts. Each 'CDS' (coding sequence) or 'mat_peptide' (mature peptide) described in the feature table was interpreted using feature locations, qualifiers and join specifications. Entries were rejected if they were near duplicates in the termination region, they had no stop codon, the stop codon was not preceded by a valid open reading frame (i.e. the sequence is not consistent with that specified in the feature table) or if the open reading frames were <100 bases. For valid coding regions, the stop codon, the next base and the following 100 bases of non-coding sequence were analysed. If the feature table described another following coding sequence, the flanking sequence was truncated to include only non-coding sequences. Then the frequency of occurrence of the 12 tetranucleotides in any frame of these non-coding regions was calculated. The group of genes having the top 10% of codon adaptation index (CAI) values were used as a subset of highly expressed genes and analysed separately; this subset is also available from the TransTerm database.

Growth media and bacterial strains

Bacteria were grown in LB medium and/or M9 medium supplemented with glucose, thiamine and necessary amino acids at recommended concentrations as described previously (Sambrook *et al.*, 1989). Bacteria containing plasmids were selected using ampicillin (100 μg/ml). Protein expression induced from the P_{tac} promotor was achieved using IPTG at a final concentration of 1 mM in the media. The *E.coli* strain TG1 (Sambrook *et al.*, 1989) was used for primary cloning and the *E.coli* strain FJU112 (Jørgensen and Kurland, 1990) was used for analysis of frameshift constructs. Strain TG1 contains a glutamine-inserting UAG suppressor tRNA which could compete with RF-1 selection in the UAGN series and was not used for comparative assays. However, strain TG1 was used to facilitate detection of recombinant clones. Strain FJU112 [Δ(lac pro) gyrA ara recA56^Tn*10*, F'lacI^{Q1}] has wild-type ribosomes and no suppressor tRNAs which could compete with termination or frameshifting events.

Plasmid construction

Complementary deoxyoligonucleotides spanning the RF-2 frameshift window containing all three UAAN, UGAN and UAGN stop signal series and the UGGN control series were annealed and cloned into the *Eco*RI and *Sal*I sites of the pMAL™ polylinker using standard recombination techniques (Sambrook *et al.*, 1989). Plasmids were electroporated (2.5 kV, 5–6 ms) into strain TG1. Cells containing plasmids were selected using ampicillin and recombinant clones detected by hybridization with one of the oligonucleotides labelled with [γ-^{32}P]dATP. Bacteria containing positive clones were screened for stop signal sequences by culturing in LB media containing IPTG. Expressed MBP fusion proteins were detected after separation by electrophoresis and staining with Coomassie brilliant blue. DNA from clones showing MPB fusion proteins was sequenced to confirm identity. DNA from separate clones of each of the 16 possible constructs was then electroporated into strain FJU112 for Western analysis of expressed proteins.

Expression and analysis of fusion proteins

For analysis of fusion proteins, an aliquot from FJU112 clones grown overnight in LB media containing ampicillin (100 μg/ml) was inoculated into 3 ml of the same media. Bacteria were grown to an A_{600} of 0.5, then IPTG added to 1 mM to induce expression from the P_{tac} promotor. After a further 2 h growth, 500 μl of each culture were removed, the bacteria pelleted by centrifugation and then lysed in 100 μl of sample buffer containing 0.01 M sodium phosphate (pH 7.2), 1% SDS, 1% β-mercaptoethanol and 6 M urea. The samples were further diluted (1:160) in the same buffer and heated to 96°C for 5 min prior to protein separation by PAGE. Proteins (15 μl of each sample) were resolved through 16% separating gels (ratio of acrylamide:bisacrylamide, 150:1) and 10% stacking gels (ratio of acrylamide:bisacrylamide, 49:1) according to the method of Kolbe *et al.* (1984). Electrophoresis was carried out at 200 V for 1 h.

Separated proteins were transferred to nitrocellulose membranes (Schleicher & Schuell) at 100 V for 1 h in a buffer that contained 25 mM Tris, 192 mM glycine (pH 8.3) and 20% methanol. The membranes were blocked overnight in TBS [40 mM Tris (pH 7.6), 150 mM NaCl, 0.05% Tween 20] containing skimmed milk powder (10 mg/ml; TBSM). The immobilized proteins were reacted with a rabbit

antibody against the MBP (1:6600 dilution in TBS, 2 h). After washing (4× 5 min with TBSM), bound antibody was tagged with an alkaline phosphatase-labelled sheep anti-rabbit antibody (Sigma; 1:5000 dilution in TBSM, 2 h). The filters were then washed (3× 5 min with TBS) and alkaline phosphatase catalysis of a substrate [83.3 μg/ml BCIP, 166.6 μg/ml NBT, in a buffer containing 100 mM Tris base (pH 9.5), 100 mM NaCl and 5 mM $CaCl_2$] allowed the ratio of the expressed proteins to be determined by laser densitometry. Expressed proteins from some constructs were tested at several concentrations to ensure that the results from densitometry of the immunostained bands were consistent over a range of dilutions.

Relative rates of RF selection

The original formula derived by Curran and Yarus (1988), and subsequently used by Pedersen and Curran (1991) to calculate the relative rate of RF-1 selection at a UAG codon, has been adapted for our experiments. Although originally this required the collection of data from separate experiments to determine the fraction of ribosomes that frameshift, in our case the data can be obtained from within the same experiment. The fraction of ribosomes that frameshift is determined from the ratio of the frameshift product to the sum of the termination and frameshifting products.

Acknowledgements

We thank S.Mannering for help in manuscript preparation. This work was supported by a grant from the Human Frontier Science Program (awarded to Yoshikazu Nakamura and W.P.T.). W.P.T. is an International Scholar of the Howard Hughes Medical Institute.

References

Adamski,F.M., Donly,B.C. and Tate,W.P. (1993) *Nucleic Acids Res.*, **21**, 5074–5078.

Adamski,F.M., McCaughan,K.K., Jørgensen,F., Kurland,C.G. and Tate,W.P. (1994) *J. Mol. Biol.*, **238**, 302–308.

Arkov,A.L., Korolev,S.V. and Kisselev,L.L. (1993) *Nucleic Acids Res.*, **21**, 2891–2897.

Atkins,J.F., Weiss,R.B. and Gesteland,R.F. (1990) *Cell*, **62**, 413–423.

Böck,A., Forchammer,K., Heider,J., Leinfelder,W., Sawers,G., Veprek,B. and Zinoni,F. (1991) *Mol. Microbiol.*, **5**, 515–520.

Bossi,L. (1983) *J. Mol. Biol.*, **164**, 73–87.

Bossi,L. and Roth,J.R. (1980) *Nature*, **286**, 123–127.

Brown,C.M., Stockwell,P.A., Trotman,C.N.A. and Tate,W.P. (1990) *Nucleic Acids Res.*, **18**, 2079–2086.

Brown,C.M., Dalpin,M.E., Stockwell,P.A. and Tate,W.P. (1993a) *Nucleic Acids Res.*, **21**, 3119–3123.

Brown,C.M., McCaughan,K.K. and Tate,W.P. (1993b) *Nucleic Acids Res.*, **21**, 2109–2115.

Brown,C.M., Stockwell,P.A., Dalphin,M.E. and Tate,W.P. (1994) *Nucleic Acids Res.*, **22**, 3620–3624.

Buckingham,R.H., Sörensen,P., Pagel,F.T., Hijazi,K.A., Mims,B.H., Brechemier-Baey,D. and Murgola,E.J. (1990) *Biochim. Biophys. Acta*, **1050**, 259–262.

Craigen,W.J. and Caskey,C.T. (1986) *Nature*, **322**, 273–275.

Craigen,W.J., Cook,R.G., Tate,W.P. and Caskey,C.T. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 3616–3620.

Craigen,W.J., Lee,C.C. and Caskey,C.T. (1990) *Mol. Microbiol.*, **4**, 861–865.

Crick,F. (1966) *Cold Spring Harbor Symp. Quant. Biol.*, **31**, 3–9.

Curran,J.F. (1993) *Nucleic Acids Res.*, **21**, 1837–1843.

Curran,J.F. and Yarus,M. (1988) *J. Mol. Biol.*, **203**, 75–83.

Curran,J.F. and Yarus,M. (1989) *J. Mol. Biol.*, **209**, 65–77.

Donly,B.C., Edgar,C.D., Adamski,F.M. and Tate,W.P. (1990) *Nucleic Acids Res.*, **18**, 6517–6522.

Eggertsson,G. and Söll,D. (1988) *Microbiol. Rev.*, **52**, 354–374.

Fluck,M.M., Salser,W. and Epstein,R.H. (1977) *Mol. Gen. Genet.*, **151**, 137–149.

Gesteland,R.F., Weiss,R.B. and Atkins,J.F. (1992) *Science*, **257**, 1640–1641.

Hatfield,D. and Oroszlan,S. (1990) *Trends Biochem. Sci.*, **15**, 186–190.

Hatfield,D.L., Smith,D.W.E., Lee,B.J., Worland,P.J. and Oroszlan,S. (1990) *Crit. Rev. Biochem. Mol. Biol.*, **25**, 71–96.

Hirsh,D. and Gold,L. (1971) *J. Mol. Biol.*, **58**, 439–458.

Jørgensen,F. and Kurland,C.G. (1990) *J. Mol. Biol.*, **215**, 511–521.

Kawakami,K., Inada,T. and Nakamura,Y. (1988) *J. Bacteriol.*, **170**, 5378–5381.

Kolbe,H.V.J., Costello,D., Wong,A., Lu,R.C. and Wohlrab,H. (1984) *J. Biol. Chem.*, **259**, 9115–9120.

Kopelowitz,J., Hampe,C., Goldman,R., Reches,M. and Engelberg-Kulka,H. (1992) *J. Mol. Biol.*, **225**, 261–269.

Marshall,B. and Levy,B. (1980) *Nature*, **286**, 524–525.

Martin,R., Weiner,M. and Gallant,J. (1988) *J. Bacteriol.*, **170**, 4714–4717.

McClelland,M. and Bhagwat,A.S. (1992) *Nature*, **355**, 595–596.

Parker,J. (1989) *Microbiol. Rev.*, **53**, 273–298.

Pedersen,W.T. and Curran,J.F. (1991) *J. Mol. Biol.*, **219**, 231–241.

Phillips,G.J., Arnold,J. and Ivarie,R. (1987) *Nucleic Acids Res.*, **15**, 2627–2638.

Prescott,C.D., Kleuvers,B. and Goringer,H.U. (1991) *Biochimie*, **73**, 1121–1129.

Purohit,P. and Stern,S. (1994) *Nature*, **370**, 659–662.

Rice,C.M., Fuchs,R.F., Higgins,D.G., Stoehr,P.J. and Cameron,G.N. (1993) *Nucleic Acids Res.*, **21**, 2967–2971.

Rydén,S.M. and Isaksson,L.A. (1984) *Mol. Gen. Genet.*, **193**, 38–45.

Salser,W. (1969) *Mol. Gen. Genet.*, **105**, 125–130.

Salser,W., Fluck,M. and Epstein,R. (1969) *Cold Spring Harbor Symp. Quant. Biol.*, **34**, 513–520.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Scolnik,E., Tompkins,R., Caskey,T. and Nirenberg,M. (1968) *Proc. Natl Acad. Sci. USA*, **61**, 768–774.

Stormo,G.D., Schneider,T.D. and Gold,L. (1986) *Nucleic Acids Res.*, **14**, 6661–6679.

Tate,W.P. and Brown,C.M. (1992) *Biochemistry*, **31**, 2443–2450.

Tate,W., Greuer,B. and Brimacombe,R. (1990) *Nucleic Acids Res.*, **18**, 6537–6544.

Weiss,R.B., Dunn,D.M., Atkins,J.F. and Gesteland,R.F. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 687–693.

Weiss,R.B., Dunn,D.M., Dahlberg,A.E., Atkins,J.F. and Gesteland,R.F. (1988) *EMBO J.*, **7**, 1503–1507.

Weiss,R.B., Dunn,D.M., Atkins,J.F. and Gesteland,R.F. (1990) *Prog. Nucleic Acids Res. Mol. Biol.*, **39**, 159–183.

Yarus,M. and Curran,J. (1992) In Hatfield,D.L., Lee,B.J. and Pirtle,R.M. (eds), *Transfer RNA in Protein Synthesis.* CRC Press, Boca Raton, FL, pp. 319–365.

Zinoni,F., Birkmann,A., Stadtman,T.C. and Böck,A. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 4650–4654.

Received on September 19, 1994